Diurnal Effects and Systematic errors in Background Determination.

Roman Fleysher   Peter Nemethy

May 22, 2001

Abstract

In this note we re-examine the grounds the standard time sloshing method is based on. Detailed analysis shows a breach induced by changes in apparatus/atmosphere: a systematic day-night variation in the angular distribution of detected showers. We find that it is necessary to modify the time sloshing method to correctly handle the modulation of zenith angle distribution. No convincing evidence is seen for a corresponding azimuth effect. The results also apply to direct integration method.

1 Motivation

In order to be able to perform a sensitive search for weak sources it is imperative to estimate background in a systematic free way. The discussion of systematic errors was started in [1]. There, problems associated with coordinate transformations were addressed. In this note we consider another possible source of a bias related to errors correlated with time and thus leading to amplification on long time scale of data processing. For that purpose we examine the grounds the time sloshing method is based on.

The basic assumption of the time sloshing method of background estimation is that a detector responds to an isotropic background. Under this supposition there should be a time independent flux of particles from each direction in local coordinates. *Thus, a shower detected with particular local coordinates could have arrived with equal probability at any other time of shower detection.*

This supposition translates to the assumption that the angular distribution of showers is time independent during the time sloshing window. Typically, time sloshing is performed over an interval of 2 hours for a point source. A search for extended sources requires the window to be enlarged. We implement an 8 hour time sloshing buffer for Galactic signal search. However, the time independence of flux in local coordinates assumption during a long time sloshing window might be violated because of changes in the atmosphere.

Changes in the atmosphere which affect shower detection are related to changes in the overburden. These can happen on daily scale due to heating and cooling of the atmosphere. One can provide a simple model of how it may happen. Let's assume that atmospheric pressure stays constant during a day, then, according to Gay-Lussac law:

$$\frac{V}{T} = const$$

or

$$nT = const$$

where n is density of molecules of air, T — temperature.

Because overburden is the integral of n over the depth of atmosphere above the detector[1], heated atmosphere will constitute less overburden than cooled one. This model suggests that during warm part of the day the detector effectively floats up to higher elevation above the sea level which should lead to a higher rate of detected showers. (Trigger threshold is lowered during this part of the day.) This change can and in fact does affect the local angle distribution of detected events which if not corrected for, will break the time sloshing assumption and introduce systematic errors in background estimation.

In the model described, zenith angle distribution of detected events is affected directly. Within the model, it seems possible that change in azimuth distribution is also induced. This may be due to atmosphere being warm "behind" the Sun and cold "in front" of it. It is also possible that solar wind modifies the Earth's magnetic field. Both and either of these may, in turn, lead to a daily modulation of the East-West effect.

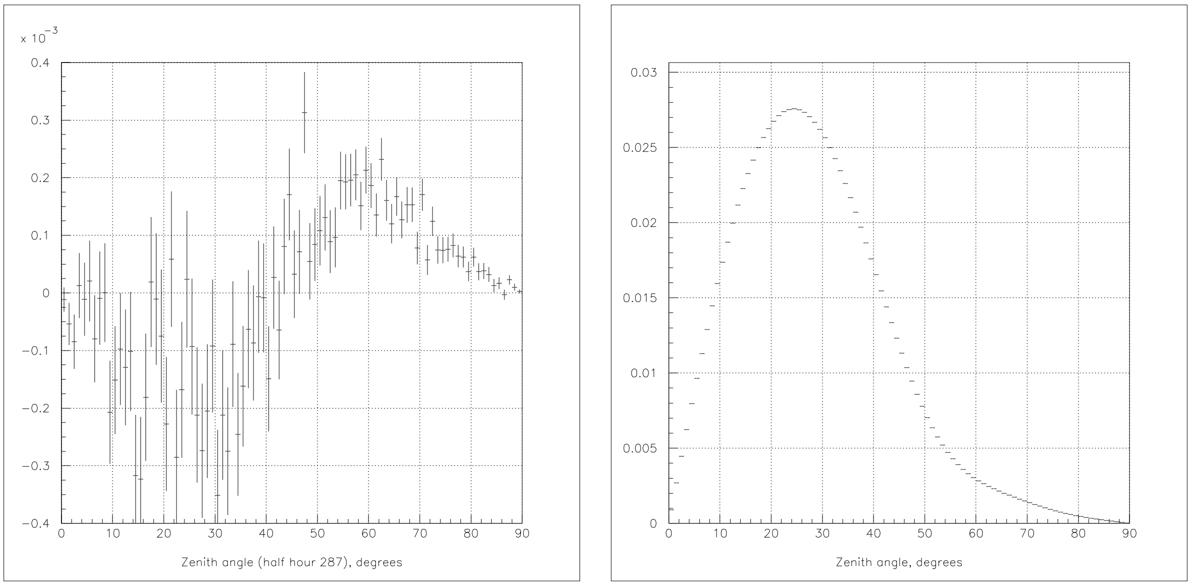
All of the above, with all of the imperfections of arguments, warrants detailed investigation of the matter.

2 Zenith distribution of events

The investigation of changes in zenith distribution was performed in the following way. Every half hour, a zenith distribution is generated from registered events. These distributions are compared to average distribution accumulated over a period of 6 days. Half hour distributions and average one are normalized and then subtracted to give a distribution's shape of change plot. An example of such a plot is presented on figure (1). A significant and smooth difference is seen.

It was observed that the shape of this change is approximately constant with amplitude varying from half hour to half hour. A plot of amplitude as a function of half hour is presented on figure (2). Event rate is also presented on the figure. These plots clearly show the diurnal property of the effect with amplitude correlating with the event rate. The correlation and periodicity provide an argument for physical nature of the phenomenon and that if different subset of data were studied, the same effect would have been observed. The amplitude plot also shows presence of changes in the zenith distribution on the time scale much longer than a day. Changes on the longer time scale than the sloshing window do not introduce problems into analysis.

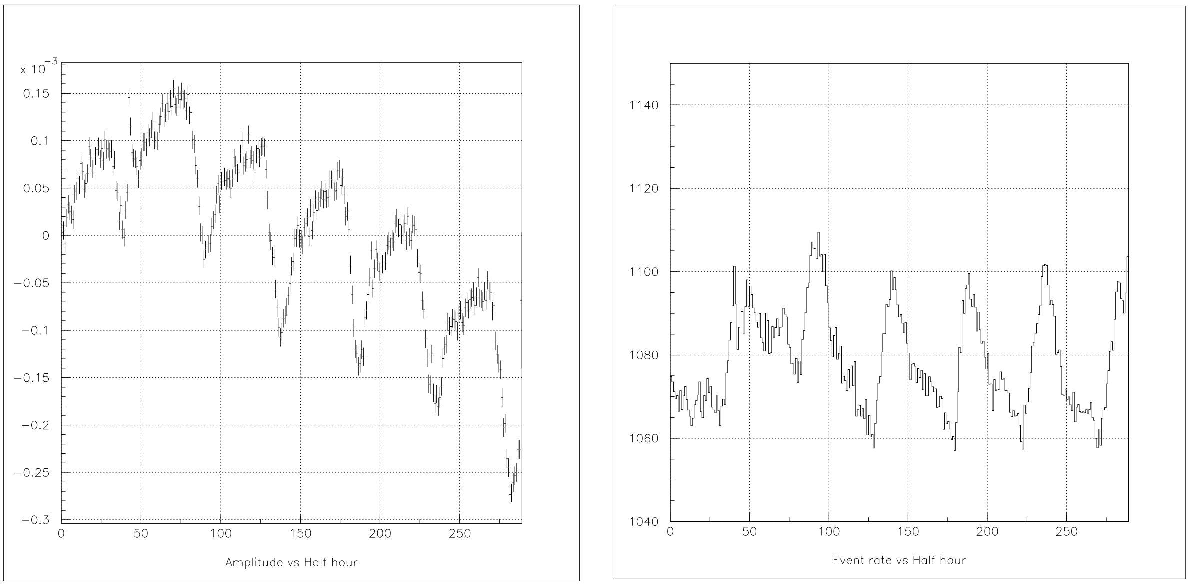
[1]It seems reasonable to assume that the depth of atmosphere (in units of length) is mostly determined by the Earth's gravitational field rather than synoptic conditions.



(a) Example of the difference of average zenith distribution and that of particular half hour.

(b) Average zenith angle distribution.

Figure 1: Plots showing average zenith angle distribution and a typical deviation from it. The fractional deviation can be as large as 1% at high zenith angles.



(a) A correction amplitude as a function of half hour.

(b) Event rate as a function of half hour.

Figure 2: Plots showing correlation between event rate and the amplitude. Left edge of the horizontal axis corresponds to beginning of a modified Julian day.

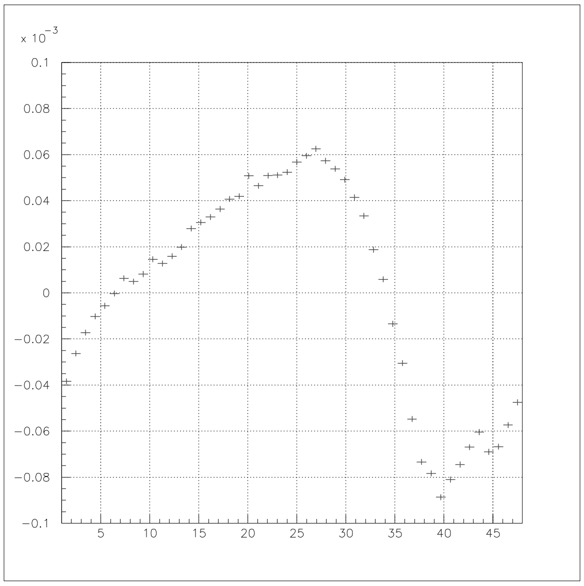


Figure 3: The correction amplitude as a function of time of day. Horizontal axis is number of half hours starting from the beginning of a modified Julian day, 6 PM local time.

3 Modification of the time sloshing method

Let the local angle distribution at time t be represented by $f(x,t)$ and average distribution over the time sloshing window by $f_0(x)$. Here x denotes the set of local coordinates of interest. Standard implementation of time sloshing will produce $f_0(x)$ as the distribution of simulated events for all times in the window. In order to incorporate time dependence one has to modify the sloshing procedure in such a way that the decision to accept the time t as a simulated arrival time of the event is based on $f(x,t)$, where x — is local coordinate of the event in question. In other words, the basic assumption of time sloshing has to be modified to become:

A shower detected with particular local coordinates x could have arrived at any other time t of shower detection with probability proportional to $f(x,t)$.

The above, for example, can be achieved with the help of the rejection method. Let $F(x)$ be a maximum of $f(x,t)$ over all t. Then, in the rejection method candidate time t is accepted as a simulated arrival time with probability $\frac{f(x,t)}{F(x)}$. If the candidate is rejected, a new one is sought randomly and then tried. It is possible to convince oneself that this modification preserves all the other standard features of the sloshing method such as event rate changes and detector down time.

4 Parametric representation

In order to be able to implement modification to the time sloshing method one has to extract the information about time dependence of the distribution function from the set of detected events. Even if one assumes that on a time scale as short as half an hour distribution does not change, Milagro registers only about $2 \cdot 10^6$ events which is not enough to determine distribution function with sufficient accuracy. Thus, a parametric representation is called for. If number of parameters is few, then $2 \cdot 10^6$ is enough to determine them with high precision.

By studying the deviations of distributions of real and simulated events (using standard sloshing method) we were able to produce such a parameterization: a polynomial correction with only one free parameter — the correction amplitude. The distribution function is considered to be of the form:

$$f(z,t) = f_0(z)e^{\theta(t)K(z)+q(\theta(t))} \tag{1}$$

where $f_0(z)$ is average zenith angle distribution as generated in standard time sloshing method, $\theta(t)$ — amplitude of the change at time t, $q(\theta(t))$ — normalization "factor". $K(z)$ — a specified function which translates to the s-like shape of figure (1(a)) Both θ and q can be easily determined from the set of real events (see appendix A) and then used in the background determination. The accuracy of the amplitude determination is shown as error bars on figure (2(a)) to be quite good.

It is interesting to note that if we plot the correction amplitude as a function of time of day (see figure 3), we observe that it follows a typical temperature behavior with one extreme at about 2 PM, the hottest time of the day, and the other at 7 AM, the coldest time, and a fast change from 7 AM to 2 PM. This is very suggestive that the change has to do with the air temperature.

We emphasize that we do not attempt to explore the correlation between amplitude, event rate, atmospheric pressure or temperature and time of the day. Instead, we let the real data determine the value of amplitude itself. We believe that the correlation can be used to establish the validity of models of the phenomenon if such are proposed rather than provide a method for amplitude estimation. Only when a model of the effect is accepted, the fact of the correlation can be exploited to supplement the measurements of the amplitude or replace them.

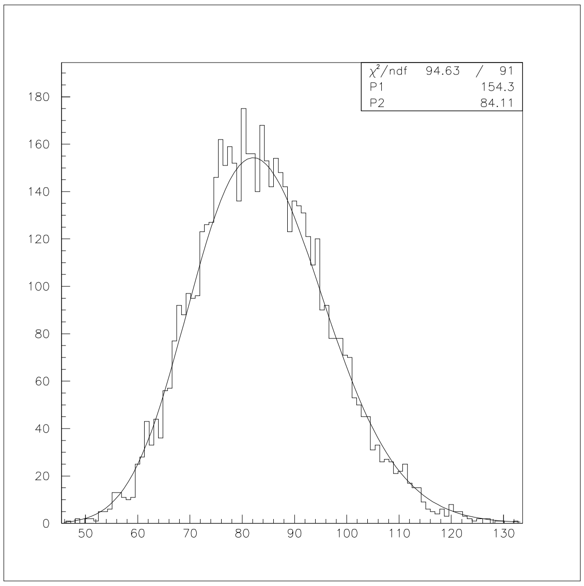
5 The correction works

In order to test how well the correction works we use χ^2 test on a sample of about 43 days of data time-sloshed 10 times. A word of precaution is appropriate before we proceed.

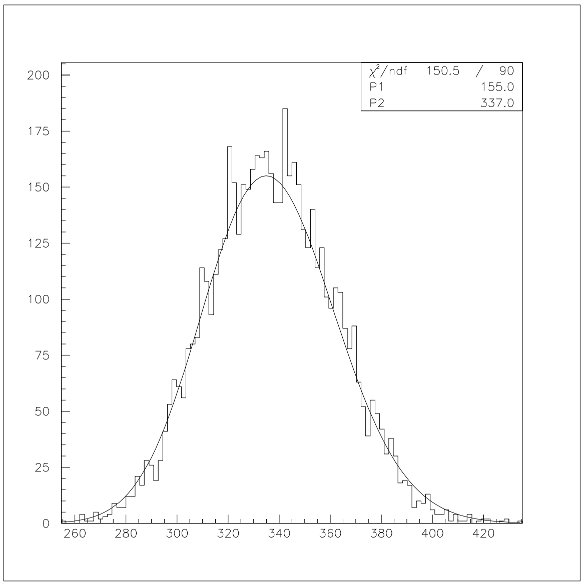
The distributions[2] we are about to compare are not independent: one is produced from the other using sloshing technique. We implement time sloshing method over the interval of 8 hours. Zenith angle distribution is histogramed every half hour, or 16 times per sloshing window. Thus, number of degrees of freedom should be lower than one would expect from the usual histogram bin considerations by 1 for every 16 bins. We employ Monte Carlo simulation of the whole process to check and support our expectations. The results of the Monte Carlo simulations are presented on figure (4) for 90 and 360 bin histograms.

In accordance to Monte Carlo simulations, 90-bin histograms should yield distribution of χ^2 with about 84.11 degrees of freedom and 360-bin ones with 337.0 degrees of freedom.

[2]Zenith angle distribution is split into 90 bins, and azimuth into 360.

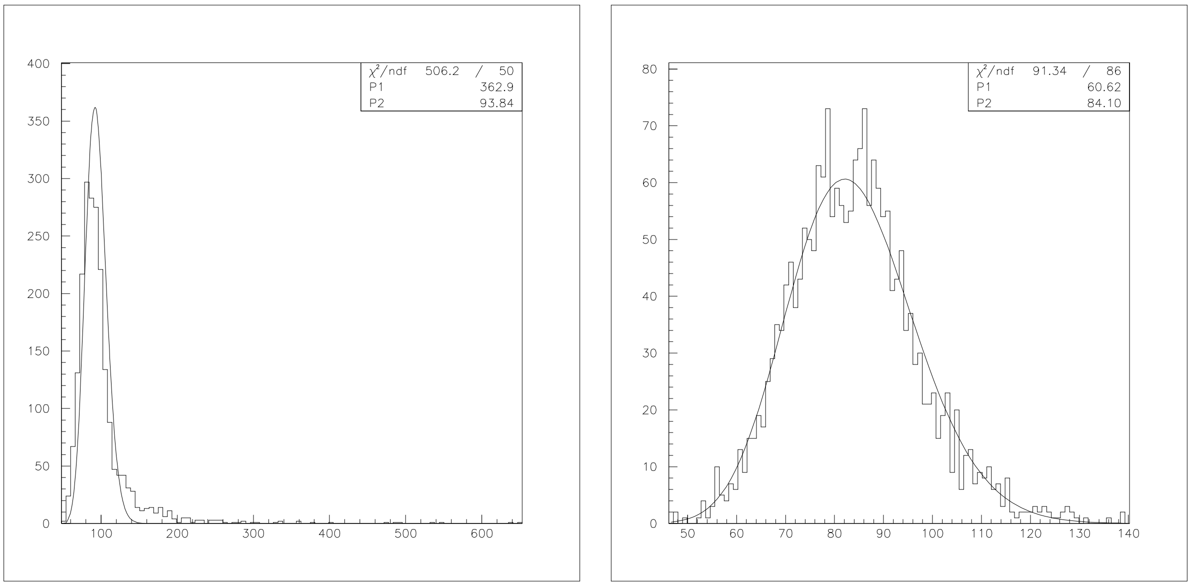


(a) Distribution of χ^2 for 90 bin histogram. Solid line is the best fit of χ^2 distribution function with 84.11 degrees of freedom.



(b) Distribution of χ^2 for 360 bin histogram. Solid line is the best fit of χ^2 distribution function with 337.0 degrees of freedom.

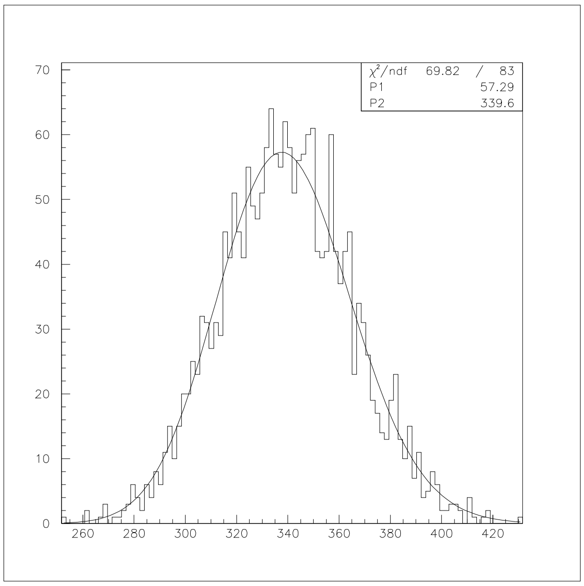
Figure 4: Plots showing Monte Carlo simulated distributions of χ^2 of the difference of histograms with 90 and 360 bins. Both histograms have the same number of entries (5713).



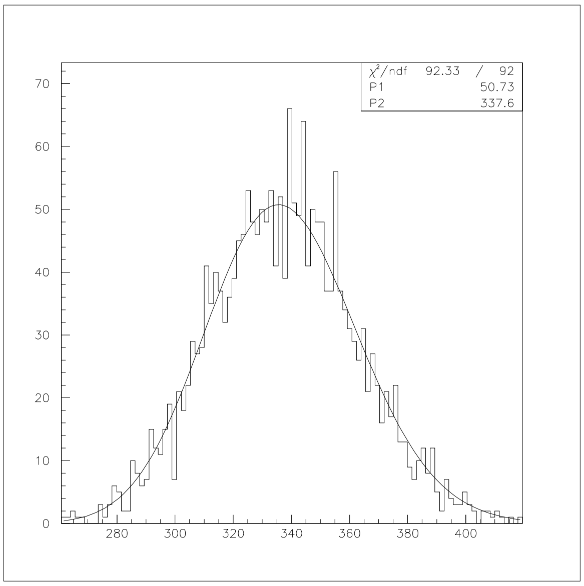
(a) Distribution of χ^2 for standard method. Solid line is the best fit of χ^2 distribution function with 93.84 degrees of freedom. The goodness of this fit is characterized by $\chi^2/ndf = 506.2/50$.

(b) Distribution of χ^2 for corrected method. Solid line is the best fit of χ^2 distribution function with 84.10 degrees of freedom. The goodness of this fit is characterized by $\chi^2/ndf = 91.34/86$.

Figure 5: Plots showing distributions of χ^2 of the difference of sloshed and real zenith angle distributions for every half hour with and without correction. The expected number of degrees of freedom is 84.25. Both histograms have the same number of entries (2073).



(a) Distribution of χ^2 for standard method. Solid line is the best fit of χ^2 distribution function with 339.6 degrees of freedom. The goodness of this fit is characterized by χ^2/ndf = 69.82/83.



(b) Distribution of χ^2 for corrected method. Solid line is the best fit of χ^2 distribution function with 337.6 degrees of freedom. The goodness of this fit is characterized by χ^2/ndf = 92.33/92.

Figure 6: Plots showing distributions of χ^2 of the difference of sloshed and real azimuth angle distributions for every half hour with and without correction. The expected number of degrees of freedom is 337.5. Both histograms have the same number of entries (2073).

Thus, Monte Carlo simulations support our expectation that we loose a factor of 1/16 in number of degrees of freedom, in which case these numbers would have been 84.25 and 337.5 respectively.

Now we are ready to look at the results of the modified sloshing method. Figure (5) demonstrates zenith angle distribution study. From these plots one can clearly see how a bias of the standard method is removed. In fact, we observed systematic problems at the 10^{-3} level in background subtracted sky maps, when using standard time sloshing with an 8 hour window.

A similar study was carried out for azimuth angle distribution. Figure (6) demonstrates the results of the study and shows that azimuth variation, if any, is on a much smaller scale than the zenith one.

6 Conclusion

In this note we continued discussion of possible systematic errors introduced by the analysis method started in [1]. We re-examined the grounds standard time sloshing method is based on and found a breach. We found and introduced modification of the time sloshing method to correct zenith angle distribution generation. Azimuth angle distribution does not show problems at the level studied. We continue our investigation of smaller azimuth and remaining uncorrected zenith effects. These results also apply to the direct integration method because it is based on the same assumptions.

7 Acknowledgement

We would like to express gratitude to Todd Haines for numerous discussions and important suggestions, which were a significant input to the preparation of this note.

A Tracking time dependence of the distribution

The choice (1) of the form of time dependence of a distribution is motivated by the Darmois theorem [2]. According to this theorem, exponential class (1) of probability density functions is the only class which admits number of sufficient statistics for unknown parameter θ independent of the number of observations. In particular, if $(Z_1, Z_2, ..., Z_N)$ is a random sample of size N from the distribution (1), then

$$Y = \sum_{i=1}^{N} K(Z_i)$$

is sufficient statistics for θ. The importance of sufficiency is explained by the fact that Y exhausts all the information about θ that is contained in the sample.

By differentiating

$$\int f(z,\theta)dz = \int f_0(z)e^{\theta K(z)+q(\theta)}dz = 1$$

twice with respect to θ we obtain the following two relations for expectation and variance:

$$E[K(Z)] = -q\prime(\theta)$$

$$var[K(Z)] = -q\prime\prime(\theta)$$

Both of these properties allow to estimate parameter θ from the data. Indeed, because $\theta = 0$ corresponds $q = 0$ (no correction to average distribution), both θ and q can be determined as

$$\begin{cases} \theta &= 2\frac{q\prime(\theta)-q\prime(0)}{q\prime\prime(\theta)+q\prime\prime(0)} + O(\theta^2) \\ q(\theta) &= \frac{1}{2}(q\prime(\theta)+q\prime(0))\theta + O(\theta^2) \end{cases}$$

Expectation and variance and therefore $q\prime$ and $q\prime\prime$ are determined from the sample and thus are known.

References

[1] Todd Haines. Systematic Errors in "Standard" Background Estimates of Milagro Data. Internal Milagro memo, 12/14/99.

[2] see for example: W.T. Eadie, D. Drijard, F.E. James, M. Ross, B. Sadoulet. Statistical Methods in Experimental Physics. North-Holland Publishing Company, 1982.